# A Matrix-Algebraic Approach to Linear Parallel Interference Cancellation in CDMA

Dongning Guo, Lars K. Rasmussen, Sumei Sun and Teng Joon Lim

*Abstract*— **In this paper linear parallel interference cancellation (PIC) schemes are described and analysed using matrix algebra. It is shown that the linear PIC, whether conventional or weighted, can be seen as a linear matrix filter applied directly to the chip-matched filtered received signal vector. An expression for the exact bit error rate is obtained and conditions on the eigenvalues of the code correlation matrix and the weighting factors to ensure convergence are derived. The close relationship between the linear multistage PIC and the steepest descent method (SDM) for minimising the mean squared error (MSE) is demonstrated. A modified weighted PIC structure that resembles the SDM is suggested which approaches the MMSE detector rather than the decorrelator. It is shown that for a $K$-user system only $K$ PIC stages are required for the equivalent matrix filter to be identical to the the MMSE filter. For fewer stages, techniques are devised for optimising the choice of weights with respect to the mean squared error. One unique optimal choice of weights is found which will lead to the minimum achievable MSE at the final stage. Simulation results show that a few stages is sufficient for near-MMSE performance.**

Index Terms—Code-division multiaccess, linear algebra, multiuser channels, signal detection.

## I. Introduction

In a mobile communications system multiple access to the common channel resources is vital. In a system based on spread-spectrum transmission techniques direct sequence (DS) code division provides simultaneous access for multiple users [1], [2]. By selecting mutually orthogonal spreading codes for all users, they each achieve interference free single-user performance. It is however not possible in a mobile environment to maintain orthogonality of the codes at the receiver and thus multiple-access interference (MAI) arises. Conventional single-user detection is severely affected by the MAI, and a system using conventional detection is interference limited [3]. Strict power control is also required to alleviate the near-far problem where a high-powered user creates overwhelming MAI for low-powered users.

Advanced multiuser detection strategies can be adopted to improve overall performance of a CDMA system [4]. In [5], Verdú developed the optimal {0,1}-constrained maximum-likelihood (ML) detector. The inherent complexity however increases exponentially with the number of users, rendering this optimal detector impractical. Two types of linear detectors have also been suggested, namely the decorrelating detector (or the decorrelator) [6] and the MMSE detector [7]. They are near-far resistant [8] and the complexity is linear in the number of users.

For practical implementation successive interference cancellation (SIC) and parallel interference cancellation (PIC) schemes have been subject to most attention. These techniques rely on simple processing elements constructed around the matched filter concept. The detector makes use of tentative decisions to estimate all interference for a certain user and purifies the decision statistic by subtracting the estimated interference before matched filtering to reach a better decision. The first structure based on the principle of interference cancellation was the parallel multistage detector proposed by Varanasi and Aazhang in [9]. A significant improvement to the PIC was suggested by Divsalar *et al.* in [10] where they proposed a weighted (or partial) cancellation scheme. They studied both linear and non-linear decisions functions based on joint ML considerations. Linear interference cancellation schemes have been further exploited by many authors. In [11], Elders-Boll *et al.* pointed out that linear SIC and PIC correspond to the Gauss-Seidel and Jacobi iteration [12], respectively, for approximating matrix inversion. The linear SIC has been analysed by Rasmussen *et al.* in [13]. In [14] Moshavi *et al.* developed a weighted matrix expansion detector which is very similar to the weighted linear PIC.

In this paper we mathematically describe linear PIC schemes using matrix algebra. Assuming a symbol synchronous system[1] using short codes, we show that the linear PIC schemes correspond to linear matrix filtering that can be performed directly on the received chip-matched filtered signal vector. The approach applies to both conventional and weighted structures. This matrix filter is referred to as the equivalent one-shot cancellation filter (a similar description has been made for linear SIC in [13]). It is then possible to get an analytical expression for the exact bit error rate (BER) and to derive conditions on the eigenvalues of the code correlation matrix and the weighting factors, or weights, to ensure convergence. The weighted linear multistage PIC is found to resemble the steepest descent method (SDM) for updating adaptive filter tap-weights to minimise the MSE. Here we demonstrate the close relationship between the two and present a new PIC structure

D. Guo, S. Sun and T. J. Lim are with the Centre for Wireless Communications, 20 Science Park Road, #02-34/37 Teletech Park, Singapore Science Park II, Singapore 117674 (email: {cwcguodn, cwcsunsm, cwclimtj}@leonis.nus.edu.sg, FAX: +65 779 5441). L. K. Rasmussen was with the Centre for Wireless Communications, Singapore. He is currently with the Department of Computer Engineering, Chalmers University of Technology, SE-412 96 Göteborg, Sweden (email: larsr@ce.chalmers.se, FAX: +46 31 772 3663). This work was supported in part by Oki Techno Centre (S'pore) Pte Ltd and the National Science and Technology Board (NSTB), Singapore. The Centre for Wireless Communications is a National Research Centre funded by NSTB.

[1] The principles apply equally well to an asynchronous system. For notational ease, we will focus on the symbol synchronous case here.

which in fact is a modified version of the structures suggested in [10] and [15]. Through multistage processing, this new structure ensures convergence to the performance of the MMSE detector rather than the decorrelating detector which other linear interference cancellation structures generally converge to. Following the principles of the SDM, we derive the corresponding one-shot cancellation filter and devise techniques for optimising the choice of weights (or equivalently the step sizes for the SDM) with respect to the mean squared error (MSE). For a $K$-user system, only $K$ PIC stages are required for the equivalent one-shot filter to be identical to the MMSE filter. For fewer stages, it is shown that one unique optimal choice of weights exists which will lead to the minimum achievable MSE at the final stage. This approach has also been applied to the case of long codes in [16].

The paper is organised as follows. In Section II, the uplink model is briefly described. The algebraic description of the conventional PIC scheme is presented in Section III together with convergence analysis. The close connection between weighted PIC and the SDM is then explained and a new structure is proposed in Section IV. In Section V the optimisation of the weights is described and powerful techniques for obtaining these parameters devised. Numerical examples are presented in Section VI and the paper is completed by some concluding remarks.

Throughout this paper scalars are lower-case, vectors are bold face lower-case, and matrices are bold face upper-case unless otherwise stated. All vectors are defined as column vectors with row vectors represented by transposition. The following notation is used for the product of matrices,

$$\prod_{i=n_1}^{n_2} \mathbf{X}_i = \begin{cases} \mathbf{X}_{n_2}\mathbf{X}_{n_2-1}\cdots\mathbf{X}_{n_1+1}\mathbf{X}_{n_1} & \text{if } n_1 \le n_2, \\ \mathbf{I} & \text{if } n_1 > n_2. \end{cases} \quad (1)$$

Subscripting is done according to the following conventions. Variables independent of the detector stage are, when needed, subscripted with a user index, e.g., $x_k$. The first subscript on variables dependent on the detector stage, e.g., $x_{i,k}$, denotes the current stage, the second subscript the user index.

## II. System Models

The discrete-time model for the uplink of a CDMA communication system considered throughout this paper is described here. It is based on a symbol-synchronous system assuming single-path channels and the presence of stationary additive white Gaussian noise (AWGN).

A specific user in this $K$-user CDMA system transmits an $M$-ary PSK information symbol $d_k \in \{\exp(j2\pi(i-1)/M)\}$, $i = 1, 2, ..., M$, by multiplying the symbol with a $q$-ary spreading code $\mathbf{s}_k \in \{\varrho_1, ..., \varrho_q\}^N$, $\varrho_j \in \mathbb{C}$ of length $N$ chips and then transmitting over an AWGN channel. All users' codes can be written concisely as an $N \times K$ matrix $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K)$ and we assume that $\mathbf{s}_k^{\mathrm{H}}\mathbf{s}_k = 1$. The received signal energy per symbol of user $k$ is denoted by $w_k$. The received chip matched filtered signal vector, $\mathbf{r}$, encompassing one symbol interval for all users, is then expressed as a weighted linear combination of the spreading codes, i.e.,

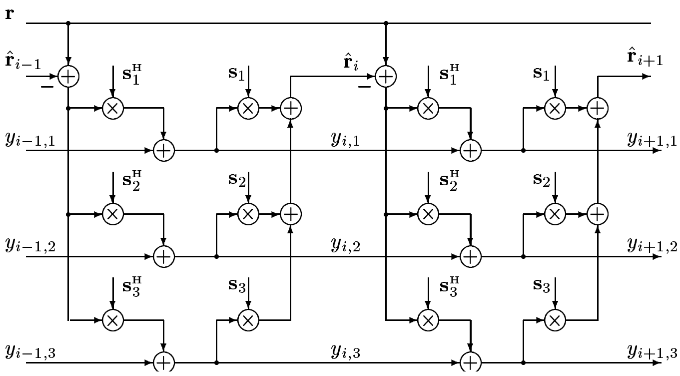


Fig. 1. The conventional PIC structure for a 3-user case. Stages $i$ and $(i+1)$ are shown. Assume that $\hat{\mathbf{r}}_0 = \mathbf{0}$ and $y_{0,k} = 0$ for all $k$'s.

$$\mathbf{r} = \sum_{k=1}^{K} \sqrt{w_k}\mathbf{s}_k d_k + \mathbf{n} = \mathbf{A}\mathbf{d} + \mathbf{n} \in \mathbb{C}^N, \quad (2)$$

where $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_K) = \left(\sqrt{w_1}\mathbf{s}_1, \sqrt{w_2}\mathbf{s}_2, \cdots, \sqrt{w_K}\mathbf{s}_K\right)$, $\mathbf{d} = (d_1, d_2, \cdots, d_K)^{\top}$, and $\mathbf{n}$ is a noise vector where each sample is independently, circularly complex Gaussian distributed with zero mean and variance $\mathsf{E}\{\mathbf{n}\mathbf{n}^{\mathrm{H}}\} = \sigma^2\mathbf{I}$.

## III. Algebraic Description of Conventional PIC

In a conventional PIC structure, previous tentative decisions are used to estimate the interference for full cancellation. The structure is depicted by the diagram in Fig. 1. The decision statistic at stage $(i+1)$ for user $k$ is then

$$y_{i+1,k} = \mathbf{s}_k^{\mathrm{H}}\left(\mathbf{r} - \sum_{\substack{j=1 \\ j\neq k}}^{K}\mathbf{s}_j\hat{d}_{i,j}\right) = \mathbf{s}_k^{\mathrm{H}}\left(\mathbf{r} - \sum_{j=1}^{K}\mathbf{s}_j\hat{d}_{i,j}\right) + \hat{d}_{i,k},$$

where $\hat{d}_{i,k}$ is a tentative decision for user $k$ at stage $i$, which is equal to $y_{i,k}$ for a linear PIC structure. Also define the vector of decision statistics at the $i^{\text{th}}$ stage as $\mathbf{y}_i = (y_{i,1}, y_{i,2}, ..., y_{i,K})^{\top}$. We then have

$$\mathbf{y}_{i+1} = \mathbf{S}^{\mathrm{H}}(\mathbf{r} - \mathbf{S}\mathbf{y}_i) + \mathbf{y}_i = \mathbf{S}^{\mathrm{H}}\mathbf{r} + (\mathbf{I} - \mathbf{S}^{\mathrm{H}}\mathbf{S})\mathbf{y}_i, \quad (3)$$

where $\mathbf{y}_0 = \mathbf{0}$. Using this recursion, we can express the soft decision output of an $m$-stage PIC as $\mathbf{y}_m = \mathbf{G}_m^{\mathrm{H}}\mathbf{r}$ where

$$\mathbf{G}_m^{\mathrm{H}} = \sum_{i=1}^{m}(\mathbf{I} - \mathbf{S}^{\mathrm{H}}\mathbf{S})^{(i-1)}\mathbf{S}^{\mathrm{H}}.$$

This is the equivalent one-shot cancellation filter for the conventional linear PIC. It is not difficult to see that all eigenvalues of $\mathbf{S}^{\mathrm{H}}\mathbf{S}$ must be less than 2 for $\mathbf{G}_m^{\mathrm{H}}$ to converge as $m$ goes to infinity. This can hardly be guaranteed for most spreading codes, hence the unstable performance of the conventional PIC (See e.g. [17]).

Since $\mathbf{G}_m^{\mathrm{H}}$ is a linear filter, the noise component in $\mathbf{y}_m$ is still Gaussian but with colored correlation matrix

$$\mathsf{E}\left\{\mathbf{G}_m^{\mathrm{H}}\mathbf{n}\mathbf{n}^{\mathrm{H}}\mathbf{G}_m\right\} = \sigma^2\mathbf{G}_m^{\mathrm{H}}\mathbf{G}_m.$$

We can therefore analytically calculate the BER for any user at any stage using the same techniques as for the conventional matched filter detector [6] and the linear successive canceller [13]. Specifically for BPSK data modulation, the BER of user $k$ at stage $m$ is

$$P_b(m,k) = \frac{1}{2^{K-1}} \sum_{\substack{\mathbf{d}\in\{-1,1\}^K \\ d_k=1}} Q\left(\frac{\Re\{\mathbf{g}_{m,k}^{\mathrm{H}}\mathbf{A}\mathbf{d}\}}{\sigma\|\Re\{\mathbf{g}_{m,k}\}\|}\right), \qquad (4)$$

where $\mathbf{g}_{m,k}$ is column $k$ of $\mathbf{G}_m$.

## IV. Steepest Descent Method vs PIC

In this section we first show the resemblance of the steepest descent method for approaching the MMSE detector to a multistage PIC. A modified multistage PIC structure is then proposed which converges to the MMSE detector. Conditions for its convergence are also discussed.

A linear detector $\mathbf{G}^{\mathrm{H}}$ is a $K \times N$ linear matrix filter applied to the chip matched received signal vector that gives the following estimate of all users' transmitted data symbols,

$$\mathbf{y} = \mathbf{G}^{\mathrm{H}}\mathbf{r} = \mathbf{G}^{\mathrm{H}}(\mathbf{A}\mathbf{d} + \mathbf{n}).$$

The corresponding MSE is a scalar defined as

$$\begin{aligned} J &= \mathsf{E}\left\{\|\mathbf{y} - \mathbf{d}\|^2\right\} = \mathsf{E}\left\{\|\mathbf{G}^{\mathrm{H}}\mathbf{r} - \mathbf{d}\|^2\right\} \\ &= \mathsf{E}\left\{\mathbf{r}^{\mathrm{H}}\mathbf{G}\mathbf{G}^{\mathrm{H}}\mathbf{r} - \mathbf{d}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{r} - \mathbf{r}^{\mathrm{H}}\mathbf{G}\mathbf{d} + \mathbf{d}^{\mathrm{H}}\mathbf{d}\right\}. \end{aligned} \qquad (5)$$

Its gradient with respect to $\mathbf{G}^{\mathrm{H}}$, assuming that $\mathsf{E}\left\{\mathbf{d}\mathbf{d}^{\mathrm{H}}\right\} = \mathbf{I}$ for independent users, is [18]

$$\nabla J = \frac{\partial J}{\partial \mathbf{G}^{\top}} = \mathbf{G}^{\mathrm{H}}\mathsf{E}\left\{\mathbf{r}\mathbf{r}^{\mathrm{H}}\right\} - \mathsf{E}\left\{\mathbf{d}\mathbf{r}^{\mathrm{H}}\right\} = \mathbf{G}^{\mathrm{H}}(\mathbf{A}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}) - \mathbf{A}^{\mathrm{H}}.$$

The steepest descent method [18] gives the following recursion for approaching the MMSE filter,

$$\mathbf{G}_i^{\mathrm{H}} = \mathbf{G}_{i-1}^{\mathrm{H}} - \mu_i\nabla J = \mathbf{G}_{i-1}^{\mathrm{H}} - \mu_i\left[\mathbf{G}_{i-1}^{\mathrm{H}}(\mathbf{A}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}) - \mathbf{A}^{\mathrm{H}}\right], \qquad (6)$$

where $\mu_i$ is a variable step size dependent on the current stage. Assuming $\mathbf{G}_0 = \mathbf{0}$ for obvious reasons, the non-recursive form is then obtained as

$$\mathbf{G}_i^{\mathrm{H}} = \sum_{l=1}^{i}\mu_l\mathbf{A}^{\mathrm{H}}\prod_{j=l+1}^{i}(\mathbf{I} - \mu_j(\mathbf{A}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I})). \qquad (7)$$

Observe that $\mathbf{A}^{\mathrm{H}}(\mathbf{I} - \mu_i(\mathbf{A}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}))\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{R}\mathbf{A}^{\mathrm{H}}(\mathbf{I} - \mu_i(\mathbf{A}\mathbf{A}^{\mathrm{H}}+\sigma^2\mathbf{I}))$ where $\mathbf{R} = \mathbf{A}^{\mathrm{H}}\mathbf{A}$, and so from (7) $\mathbf{G}_i^{\mathrm{H}}\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{R}\mathbf{G}_i^{\mathrm{H}}$. Substituting this expression into (6) we get

$$\mathbf{G}_i^{\mathrm{H}} = \left[\mathbf{I} - \mu_i(\mathbf{R} + \sigma^2\mathbf{I})\right]\mathbf{G}_{i-1}^{\mathrm{H}} + \mu_i\mathbf{A}^{\mathrm{H}}. \qquad (8)$$

Post-multiplying with $\mathbf{r}$ gives the update equation of the decision statistic vector,

$$\mathbf{y}_i = (\mathbf{I} - \mu_i(\mathbf{R} + \sigma^2\mathbf{I}))\mathbf{y}_{i-1} + \mu_i\mathbf{A}^{\mathrm{H}}\mathbf{r}. \qquad (9)$$

The MMSE detector implemented using this steepest descent updating equation may be seen as a new parallel interference cancellation structure as illustrated in Fig. 2 [19]. Here we introduce a non-negative real parameter $\alpha$ in replace of $\sigma^2$, so that the update is expressed as

$$\mathbf{y}_i = (\mathbf{I} - \mu_i(\mathbf{R} + \alpha\mathbf{I}))\mathbf{y}_{i-1} + \mu_i\mathbf{A}^{\mathrm{H}}\mathbf{r}. \qquad (10)$$

When $\alpha$ is set to $\sigma^2$, the structure implements the steepest descent method exactly. The reason for introducing this $\alpha$ is because $\sigma^2$ is in general not known at the receiver. The effect of $\alpha$ will be discussed later.

Interestingly, this new structure can be customised to most other linear PIC detectors that have been suggested. For $\mu_i = 1$, $\alpha = 0$ and equal power users[2], Eqns. (10) and (3) are identical, hence the structure may be tuned to the conventional PIC. Also assuming $\alpha = 0$ and equal power users but varying $\mu_i$, (10) describes the structure in [10]. For a fixed $\mu_i = \mu$, (10) describes the structure in [15]. It is also obvious that a 1-stage PIC is equivalent to the conventional single-user detector.

Based on the update equation (10), a non-recursive expression for the soft detector output of an $m$-stage PIC can be obtained as

$$\mathbf{y}_m = \left(\mathbf{I} - \prod_{i=1}^{m}(\mathbf{I} - \mu_i(\mathbf{R} + \alpha\mathbf{I}))\right)(\mathbf{R} + \alpha\mathbf{I})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{r} = \mathbf{G}_m^{\mathrm{H}}\mathbf{r}. \qquad (11)$$

It is clearly a linear matrix filtering of the received signal $\mathbf{r}$. The BER performance can then be determined analytically using similar techniques as (4).

We may also write the detector output in terms of a "steady-state" solution corrupted by some disturbance, i.e., $\mathbf{y}_m = \mathbf{y}_\infty - \mathbf{e}_m$ where $\mathbf{y}_\infty$ is the "steady-state" filter output and $\mathbf{e}_m$ is some excess transient error related to stage $m$,

$$\mathbf{e}_m = \left(\prod_{i=1}^{m}(\mathbf{I} - \mu_i(\mathbf{R} + \alpha\mathbf{I}))\right)(\mathbf{R} + \alpha\mathbf{I})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{r}. \qquad (12)$$

Based on (12) it is clear that a sufficient (but not necessary) condition[3] for convergence is,

$$-1 < 1 - \mu_i(\lambda_k + \alpha) < 1 \text{ for all } k\text{'s} \quad \Leftrightarrow \quad 0 < \mu_i < \frac{2}{\lambda_{\max} + \alpha},$$

where $\lambda_1, \lambda_2, \cdots, \lambda_K$ are the eigenvalues of the Hermitian positive semi-definite matrix $\mathbf{R}$, and $\lambda_{\max}$ denotes the largest one. The "steady-state" filter in this case would be:

$$\mathbf{G}_\infty^{\mathrm{H}} = (\mathbf{R} + \alpha\mathbf{I})^{-1}\mathbf{A}^{\mathrm{H}}. \qquad (13)$$

[2] All users have the same SNR at the receiver. $\mathbf{A} = \mathbf{S}$ and hence $\mathbf{R} = \mathbf{S}^{\mathrm{H}}\mathbf{S}$.

[3] For an identical weight to be used in all stages ($\mu_i = \mu$), it is also a necessary condition.

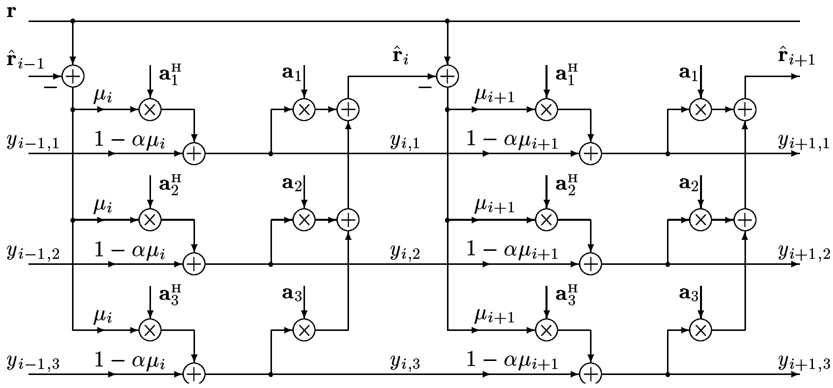


Fig. 2. The modified weighted PIC structure for a 3-user case. Stages $i$ and $(i+1)$ are shown. Assume that $\hat{\mathbf{r}}_0 = \mathbf{0}$ and $y_{0,k} = 0$ for all $k$'s.

Therefore, given a sufficiently large number of stages and an appropriate choice of weights, the "steady-state" filter is identical to the decorrelator when $\alpha = 0$, and identical to the MMSE detector when $\alpha = \sigma^2$. For the conventional PIC, $\alpha = 0$, $\mu = 1$ and $\lambda_1, \lambda_2, \cdots, \lambda_K$ are the eigenvalues of $\mathbf{S}^{\mathrm{H}}\mathbf{S}$. In this case we have convergence to the decorrelator if and only if $\lambda_{\max} < 2$ which is not true in general. As mentioned in the previous section, this accounts for the unstable behaviour of the conventional PIC.

## V. Optimisation of The Weights

In this section we consider the problem of determining a good set of weights for the PIC given that we only have a few stages ($m \ll \infty$). Our approach is to seek the optimal set of weights according to the MSE criterion. This is a natural choice since the PIC is designed as an analogy of the SDM which reduces the MSE stage by stage. Furthermore, the MSE is a quadratic function of the filter tap-weights and hence leads to a tractable optimisation problem. In addition, the MSE criterion has been verified to lead close to the minimum BER for linear detectors [20] and hence is a good choice. For notational simplicity, we first assume that $\alpha = \sigma^2$ and derive the optimal weights. The effect of setting $\alpha$ to some other value is accounted for later on.

As shown in Appendix A, the MSE of the PIC output at stage $m$ is obtained as

$$\begin{aligned} J^{(m)}(\boldsymbol{\mu}, \alpha) &= J_{\mathrm{MMSE}} + J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}, \alpha) \\ &= \sum_{k=1}^{K} \frac{\sigma^2}{\lambda_k + \sigma^2} + \sum_{k=1}^{K} \frac{\lambda_k(\lambda_k+\sigma^2)}{(\lambda_k+\alpha)^2} \\ &\quad \cdot \left| \frac{\sigma^2 - \alpha}{\lambda_k + \sigma^2} - \prod_{i=1}^{m} (1 - \mu_i(\lambda_k + \alpha)) \right|^2, \end{aligned} \tag{14}$$

where $\boldsymbol{\mu} \triangleq (\mu_1, \mu_2, ..., \mu_m)^{\top}$. The first term on the right hand side is the MMSE, which is achieved only by the MMSE filter, while the second term, called the excess MSE hereafter, represents the degradation with respect to the MMSE filter.

In [17] we considered the simple case where $\alpha = \sigma^2$ and an identical real weight, $\mu$, is used for all stages. The excess MSE is reduced to a $2m^{\mathrm{th}}$ order polynomial,

$$J_{\mathrm{ex}}^{(m)}(\mu) = \sum_{k=1}^{K} \frac{\lambda_k}{\lambda_k + \sigma^2} (1 - \mu(\lambda_k + \sigma^2))^{2m}.$$

The global minimum can be easily found as a root of the derivative polynomial. This results in a $(2m-1)^{\mathrm{th}}$ order polynomial and $(2m-1)$ potential solutions arise. However, the reader may easily verify that the second order derivative of $J_{\mathrm{ex}}^{(m)}(\mu)$ is always positive and therefore a unique real solution exists, which is the desired real-valued weight. This also enables various searching methods to be employed in locating this weight.

In the following we consider a more general case of using different weights in each stage, while assuming that $\alpha = \sigma^2$. Following Eqn. (14) we have $J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}) \triangleq J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}, \alpha = \sigma^2)$ expressed as

$$J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}) = \sum_{k=1}^{K} \frac{\lambda_k}{(\lambda_k + \sigma^2)} \prod_{i=1}^{m} |1 - \mu_i(\lambda_k + \sigma^2)|^2. \tag{15}$$

Assuming that $m \geq K$ and letting $\phi_k \triangleq \lambda_k + \sigma^2$, we can write out (15) as

$$\begin{aligned} &J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}) = \\ &\frac{\lambda_1}{\phi_1} \underline{|1-\mu_1\phi_1}|^2 |1-\mu_2\phi_1|^2 \cdots |1-\mu_K\phi_1|^2 \cdots |1-\mu_m\phi_1|^2 \\ &+ \frac{\lambda_2}{\phi_2} |1-\mu_1\phi_2|^2 \underline{|1-\mu_2\phi_2}|^2 \cdots |1-\mu_K\phi_2|^2 \cdots |1-\mu_m\phi_2|^2 \\ &\qquad \vdots \qquad\qquad \cdots \qquad\qquad \vdots \\ &+ \frac{\lambda_K}{\phi_K} |1-\mu_1\phi_K|^2 |1-\mu_2\phi_K|^2 \cdots \underline{|1-\mu_K\phi_K}|^2 \cdots |1-\mu_m\phi_K|^2. \end{aligned}$$

Obviously we can make $J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu})$ zero by selecting the weights in such a way that all the above underlined terms are zero. It is therefore clear that the MMSE solution can be achieved if

$$\mu_i = \frac{1}{\lambda_i + \sigma^2}, \quad i = 1, 2, ...K. \tag{16}$$

So the linear PIC needs at most $K$ stages to implement the MMSE detector.

In practice, however, the number of PIC stages is usually less than $K$ so we do not have enough weights to make the excess MSE zero. The objective is then to seek the global minimum of $J_{\text{ex}}^{(m)}(\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$, subject to a given number of stages $m < K$.

Note that the individual summands in (15) can be rewritten in the following way:

$$\frac{\lambda_k}{\phi_k}\prod_{i=1}^{m}|1-\mu_i\phi_k|^2 = \frac{\lambda_k}{\phi_k}\left|1+\phi_k x_1+\phi_k^2 x_2+\cdots+\phi_k^m x_m\right|^2, \tag{17}$$

where

$$\begin{aligned} x_1 &\triangleq (-1)(\mu_1+\mu_2+\cdots+\mu_m) \\ x_2 &\triangleq (-1)^2(\mu_1\mu_2+\mu_1\mu_3+\cdots+\mu_{m-1}\mu_m) \\ &\vdots \qquad\qquad \cdots \qquad\qquad \vdots \\ x_m &\triangleq (-1)^m \mu_1\mu_2\cdots\mu_m. \end{aligned} \tag{18}$$

This corresponds to the elementary symmetric polynomial transform defined as follows.

*Definition 1:* Mapping $\mathbf{T}: \mathbb{C}^m \to \mathbb{C}^m$ is given by $\mathbf{x} = \mathbf{T}(\boldsymbol{\mu})$, where $\mathbf{x} = (x_1, x_2, \cdots, x_m)^\top$ and for $i = 1, 2, \cdots, m$,

$$x_i = (-1)^i \sum_{1\le j_1<j_2<\cdots<j_i\le m} \mu_{j_1}\mu_{j_2}\cdots\mu_{j_i}. \tag{19}$$

This mapping is not one-to-one since we have in general $m!$ permutations of the same set of $\mu_1, \mu_2, \cdots, \mu_m$ for which $\mathbf{x} = \mathbf{T}(\boldsymbol{\mu})$. Based on (17), we can rewrite (15) as a quadratic function of $\mathbf{x}$ with a slight abuse of notation,

$$J_{\text{ex}}^{(m)}(\mathbf{x}) = \sum_{k=1}^{K}\frac{\lambda_k}{\phi_k}\left|1+\sum_{i=1}^{m}\phi_k^i x_i\right|^2 = \sum_{k=1}^{K}\lambda_k\phi_k\left|\frac{1}{\phi_k}+\boldsymbol{\varphi}_k^\top\mathbf{x}\right|^2,$$

where $\boldsymbol{\varphi}_k \triangleq \left(1, \phi_k, \cdots, \phi_k^{m-1}\right)^\top$. Differentiating with respect to $\mathbf{x}$ and equating to zero yields the minimum in $\mathbf{x}$,

$$\begin{aligned} \frac{\partial J_{\text{ex}}^{(m)}(\mathbf{x})}{\partial \mathbf{x}} &= \sum_{k=1}^{K}\lambda_k\phi_k\boldsymbol{\varphi}_k\boldsymbol{\varphi}_k^\top\mathbf{x} + \sum_{k=1}^{K}\lambda_k\boldsymbol{\varphi}_k = 0 \\ \Leftrightarrow &\quad \sum_{k=1}^{K}\lambda_k\phi_k\boldsymbol{\varphi}_k\boldsymbol{\varphi}_k^\top\mathbf{x} = -\sum_{k=1}^{K}\lambda_k\boldsymbol{\varphi}_k \\ \Leftrightarrow &\quad \mathbf{C}\mathbf{x} = -\mathbf{p}. \end{aligned} \tag{20}$$

Clearly, $\mathbf{C}$ is positive semi-definite since for any $\mathbf{z} \in \mathbb{R}^m$, $\mathbf{z}^\top\mathbf{C}\mathbf{z} = \sum_{k=1}^{K}\lambda_k\phi_k(\boldsymbol{\varphi}_k^\top\mathbf{z})^2 \ge 0$. Furthermore, it is shown in Appendix B that $\mathbf{C}$ is positive definite if and only if $m$ or more of the eigenvalues of $\mathbf{R}$ are distinct ($\lambda_i \ne \lambda_j$). Otherwise, $\mathbf{C}$ is singular and multiple solutions to (20) arise. Assuming that $\mathbf{C}$ is non-singular throughout this paper, which is almost always true in our cases of interest, we have a unique solution as $\hat{\mathbf{x}} = -\mathbf{C}^{-1}\mathbf{p}$, which is obviously real-valued since $\phi_k$ and $\lambda_k$ are real.

Recall that the corresponding minimum in $\boldsymbol{\mu}$ (i.e., the set of optimal weights) satisfies $\mathbf{T}(\boldsymbol{\mu}) = \hat{\mathbf{x}}$. Consider the following $m^{\text{th}}$-order polynomial with $\hat{\mathbf{x}}$ as coefficients,

$$\begin{aligned} p(\mu) &= (\mu-\mu_1)(\mu-\mu_2)\cdots(\mu-\mu_m) \\ &= \mu^m + \hat{x}_1\mu^{m-1} + \hat{x}_2\mu^{m-2} + \cdots + \hat{x}_m. \end{aligned} \tag{21}$$

It has $m$ roots, which form exactly the set of optimal weights. The reason is that the solution to this polynomial, disregarding the ordering of the roots, essentially corresponds to the inverse of the symmetric polynomial transform. Since the optimal weights are the roots of a real coefficient polynomial, they are either real or in complex conjugate pairs.

Now that we have $m$ weights, we have $m!$ different choices of $\hat{\boldsymbol{\mu}}$ as permutations of the weights that all lead to the same $J_{\text{ex}}^{(m)}(\hat{\boldsymbol{\mu}})$ at the final stage. The order in which the $m$ weights are applied however, has a significant influence on the MSE performance at intermediate stages. For a practical system, the intermediate decisions may be important for channel estimation. Depending on the desired behaviour for intermediate stages, different criteria for weight ordering can be adopted. In this paper we choose to arrange the weights according to a recursive minimisation of the excess MSE stage-by-stage. Such an ordering is obtained by selecting at stage $i$ the weight $\mu_i \in M_i$ which is closest to $\tilde{\mu}_i$, where $M_i$ denotes the set of $(m-i+1)$ elements of $\hat{\boldsymbol{\mu}}$ which have not been used in the first $(i-1)$ stages, and

$$\tilde{\mu}_i = \arg\min_{\mu_i} J_{\text{ex}}^{(i)}(\boldsymbol{\mu}_i) \tag{22}$$

where $\boldsymbol{\mu}_i = (\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_{i-1}, \mu_i)^\top$. The closest remaining weight to $\tilde{\mu}_i$ is the best choice since $J_{\text{ex}}^{(i)}(\boldsymbol{\mu}_i)$, given that all previous $\hat{\mu}_j$ is already chosen, is a quadratic function in $\mu_i$.

So far in this section, we have assumed that $\alpha = \sigma^2$ and based all derivations on (15). The derivations presented here can also be done based on (14) for an arbitrary $\alpha$. Interested readers are referred to [21] for a detailed derivation. In this case, the global minimum can be found similarly to (20). It is a function of $\alpha$, $\hat{\mathbf{x}}(\alpha)$, where the derivative of $J_{\text{ex}}^{(m)}(\mathbf{x}(\alpha), \alpha)$ with respect to $x_i$ is zero for all $i$, i.e.,

$$\frac{\partial J_{\text{ex}}^{(m)}(\mathbf{x}(\alpha), \alpha)}{\partial x_i} = 0, \quad i = 1, 2, \cdots, m. \tag{23}$$

Interestingly, the parameter $\alpha$ has no effect on the minimum excess MSE as stated in the following theorem.

*Theorem 1:* The minimum excess MSE, $J_{\text{ex}}^{(m)}(\hat{\mathbf{x}}(\alpha), \alpha)$, is independent of $\alpha$.

*Proof:* It is not difficult to show that

$$\frac{\partial J_{\text{ex}}^{(m)}(\mathbf{x}, \alpha)}{\partial \alpha} = \sum_{j=1}^{m-1} j x_{j+1}\frac{\partial J_{\text{ex}}^{(m)}(\mathbf{x}, \alpha)}{\partial x_j} \tag{24}$$

holds for all $\mathbf{x}$ and $\alpha$. At the global minimum, $\hat{\mathbf{x}}(\alpha)$, the derivative terms on the right hand side are all zero from Eqn. (23). Hence $\mathrm{d}J_{\text{ex}}^{(m)}(\hat{\mathbf{x}}(\alpha), \alpha)/\mathrm{d}\alpha = 0$. ∎

So, the specific value of $\alpha$ has no influence on the minimum achievable MSE. For any $\alpha$ we can find an $\hat{\mathbf{x}}(\alpha)$ and hence a corresponding set of weights, $\hat{\boldsymbol{\mu}}(\alpha)$, that will give us the MMSE for an $m$-stage PIC. In Appendix C, Theorem 2 states that for $\alpha = \sigma^2$ it is always true that the corresponding weights are real. For $\alpha \neq \sigma^2$ it is in general not true. As briefly mentioned earlier, the optimal weights can now be complex numbers. This is not a serious problem since the structure in Fig. 2 can easily accommodate any complex weights, although complex processing increases complexity.

If the optimal weights are found to be complex numbers, but for implementation simplicity only real parameters are wanted, they can be determined as a solution to a group of $m$ real multi-variate polynomials,

$$\frac{\partial J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}, \alpha)}{\partial \mu_j} = 0, \quad j = 1, 2, \cdots, m, \tag{25}$$

each with a degree of $(2m-1)$. Multiple solutions may exist. It is also found that the optimal set of weights should be located in the constrained set $\mathbb{B}_{\boldsymbol{\mu}} \stackrel{\triangle}{=} \{ \boldsymbol{\mu} \mid \exists\, 1 \leq i < j \leq m,\ \mu_i = \mu_j,\ \boldsymbol{\mu} \in \mathbb{R}^m \}$, i.e., at least two out of the $m$ weights are equal. This is formally proved as Theorem 3 in Appendix D. In fact $\mathbb{B}_{\boldsymbol{\mu}}$ is the set of all $\boldsymbol{\mu}$ described by the limit where potential complex conjugate pairs become real. Since there is less freedom in selecting the weights (a pair of them are forced to be equal), there will be some MSE degradation in comparison to allowing complex parameters. It is possible to locate the minimum through various searching methods such as Newton's method, but this is unfortunately too complicated to be practical.

Regardless of $\alpha$, the optimal weights, whether real or complex, are dependent on $\sigma^2$. The sensitivity of the performance of the PIC to a mismatch in $\sigma^2$ is investigated through simulations in Section VI.

In a system using long codes, the optimal weights dependent on instantaneous eigenvalues of the channel correlation matrix will have to be recomputed for every symbol interval. This is not very practical, so instead a set of time-invariant weights that minimises the ensemble average of the MSE over random codes can be used. This problem is investigated in detail in [16]

## VI. Numerical Results

The numerical examples considered in this section are based on a symbol-synchronous system with 15 users. BPSK modulation and spreading formats are assumed where short codes with a processing gain of 31 are considered. When the performance of the detector is illustrated as a function of the number of PIC stages, the SNR is 7 dB for all users.

Fig. 3 shows the performance of PIC with optimised weights. A randomly selected set of short codes is used where the corresponding correlation matrix has the eigenvalues (0.152, 0.220, 0.267, 0.350, 0.579, 0.633, 0.773, 0.884, 0.933, 1.17, 1.45, 1.55, 1.65, 1.82, 2.57). Since the largest eigenvalue is greater than 2, the condition for convergence for the conventional PIC is violated and in Fig. 3 we observe divergence. For the right choice of weights however, it is clear that a 15-stage PIC can achieve exactly the MMSE performance no matter which $\alpha$ is chosen in the structure. A 5-stage PIC gives close to MMSE performance and even a 3-stage PIC shows considerable improvement over the conventional detector[4]. The weights are ordered according to the criterion described in Section V, forcing the MSE to decrease the most, stage-by-stage. It is observed that the BER decreases monotonically when $\alpha = \sigma^2$, but not necessarily so for $\alpha \neq \sigma^2$. When the weights are not arranged properly, the BER performance can be very poor for all but the first and the last stage. This is illustrated in Fig. 3 through an example of a 5-stage PIC with $\alpha = \sigma^2$, of which the weights are sorted in decreasing order. Obviously the ordering of the weights is vital for performance in the intermediate stages.

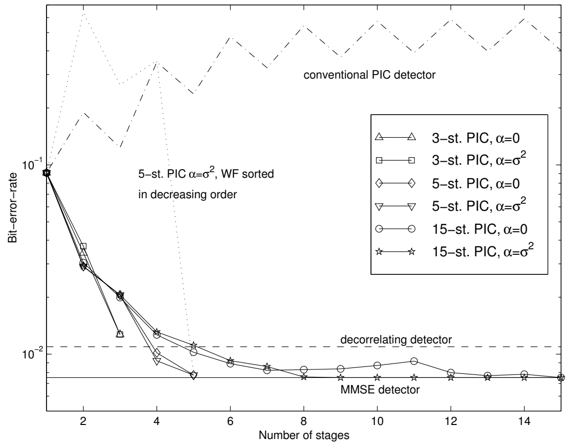


Fig. 3. Stage-by-stage performance of PIC detector with optimal weighting factors (WF) using short codes.

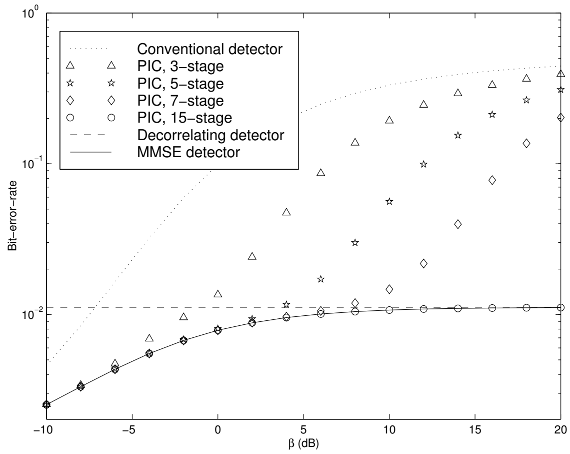


Fig. 4. Near-far ability of the weighted PIC using short codes.

Fig. 4 demonstrates the ability of a PIC to combat near-far effects. The SNR of the first user is 7 dB while the remaining 14 users have the same SNR of 7+$\beta$ dB. The weighted PIC is shown to behave better than the conventional detector in a near-far environment. The more num-

[4]The performance of the conventional detector is identical to the PIC performance in the first stage.

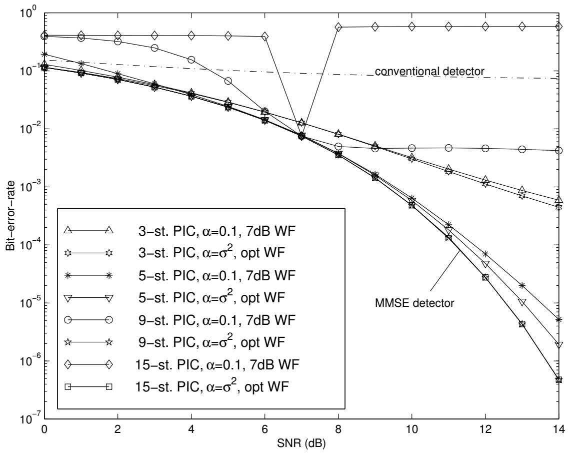


Fig. 5. BER performance and sensitivity vs. SNR using short codes. Weighting factors (WF) optimised for 7 dB and $\alpha = 0.0998$ (simplified to $\alpha = 0.1$ in the legend) are used for 0-14 dB, in comparison to when the weights are optimised for each working SNR and corresponding $\alpha$.

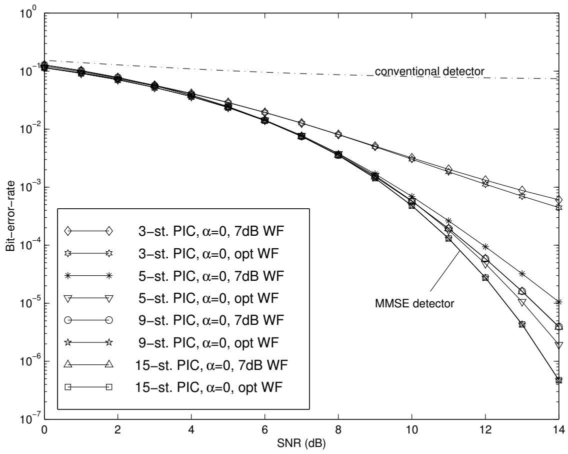


Fig. 6. BER performance and sensitivity vs. SNR using short codes. Weights (WF) for 7 dB are used for 0-14 dB, in comparison to when the weights are optimised for each working SNR. $\alpha = 0$ is assumed in all cases.

ber of stage deployed, the higher the MAI a PIC can resist. Since a 15-stage weighted PIC achieves the MMSE detector exactly, it is at least as good as the decorrelating detector in combating near-far effect. However a PIC with fewer stages is not near-far resistant in the sense that when the MAI becomes extremely strong, i.e., $\beta = \infty$, the detector fails to detect the signal from the desired user.

The weights are determined based on a specific working SNR. The sensitivity of the detector performance at all SNRs (0-14 dB), to the choice of the working SNR is illustrated in Fig. 5 and 6. The same set of short codes as for Fig. 3 is used. In Fig. 5 the weights optimised for SNR of 7 dB and $\alpha = 0.0998$ (which corresponds to a $\sigma^2$ at 7 dB) are used for various actual SNRs from 0 to 14 dB. It is compared to the case when the weights are optimised and $\alpha$ chosen for the encountered SNR under which the system is supposed to be working. The BER performance of the MMSE detector is also shown. Similar tests are done in Fig. 6, where $\alpha$ is assumed to be 0. The system is observed to be practically insensitive to SNR variation when $\alpha$ is chosen to be 0. The set of weights optimised for an SNR of 7 dB can virtually be used for any working SNR. When $\alpha = \sigma^2$ is used, the sensitivity increases substantially when the number of stages increases beyond 5. A PIC detector with more than 9 stages and 7 dB weights will perform poorly for any working SNR other than 7 dB. An intuitive explanation is provided as follows. The weights are dependent on the choice of $\alpha$ and $\sigma^2$ and found at the minimum of the excess MSE where the derivatives with respect to the weights as well as $\alpha$ are zero. If $\alpha$ is set to zero[5], then the MSE is relatively stable to a mismatch in the SNR, since $\alpha$ is fixed, i.e., only the mismatch in $\sigma^2$ affects the weights. If $\alpha = \sigma^2$, then both parameters are affected by the mismatch and thus the weights are carried further from their optimal values. As every 3 dB difference in the SNR corresponds to an additional factor of 2 in mismatch, the effect increases with increasing mismatch. This result implies that $\alpha = 0$ instead of $\alpha = \sigma^2$ should be used in a short-code PIC detector with a large number of stages. Using $\alpha = 0$ obviates estimation of the noise level but on the other hand increases detector complexity since the weights would in general be complex.

## VII. Concluding Remarks

In this paper, we have developed a matrix algebraic approach to linear parallel interference cancellation schemes. It is shown that both the conventional and the weighted linear PIC are equivalent to a one-shot linear matrix filtering. A new weighted PIC structure is suggested which approaches the performance of the MMSE detector with correctly selected weights.

Weight optimisation of the PIC is studied based on the MSE criterion. It is shown that for short-code systems, only $K$ (the number of users) stages are necessary for the PIC to be identical to the MMSE detector. For a smaller number of stages, an analytical approach is derived for finding the optimal weights that will obtain the achievable MMSE. An ordering of the weights which will ensure the largest decrease in the MSE, stage by stage, is suggested and shown to provide a monotonically decreasing BER for $\alpha = \sigma^2$. The optimal weights are dependent on the assumed SNR. It is demonstrated that for $\alpha = 0$, the detector performance is insensitive to deviation from the true SNR. For $\alpha = \sigma^2$, the detector is however, quite sensitive when a large number of stages are used.

## Appendices

### A. Derivation of the MSE

The MSE at the output of an $m$-stage PIC is

$$\begin{aligned} J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}, \alpha) &= \mathsf{E}\left\{\|\mathbf{G}_m^{\mathrm{H}}\mathbf{r} - d\|^2\right\} \\ &= \mathsf{E}\left\{\mathbf{r}^{\mathrm{H}}\mathbf{G}_m\mathbf{G}_m^{\mathrm{H}}\mathbf{r} - 2\Re\{\mathbf{d}^{\mathrm{H}}\mathbf{G}_m^{\mathrm{H}}\mathbf{r}\} + \mathbf{d}^{\mathrm{H}}\mathbf{d}\right\}, \end{aligned}$$

[5]Or any other fixed value.

where $\mathbf{G}_m^{\mathrm{H}}$ is described by (11) and $\Re\{\cdot\}$ stands for the real part of a complex number or matrix. Since $J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu},\alpha)$ is a scalar, we can apply the trace operator to the right hand side directly. Utilizing the fact that $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$, we have

$$\begin{aligned}
&J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu},\alpha)\\
=&\mathrm{tr}\left\{\mathsf{E}\left\{\mathbf{r}^{\mathrm{H}}\mathbf{G}_m\mathbf{G}_m^{\mathrm{H}}\mathbf{r} - 2\Re\{\mathbf{d}^{\mathrm{H}}\mathbf{G}_m^{\mathrm{H}}\mathbf{r}\} + \mathbf{d}^{\mathrm{H}}\mathbf{d}\right\}\right\}\\
=&\mathrm{tr}\left\{\mathbf{G}_m^{\mathrm{H}}\mathsf{E}\left\{\mathbf{r}\mathbf{r}^{\mathrm{H}}\right\}\mathbf{G}_m - 2\Re\{\mathbf{G}_m^{\mathrm{H}}\mathsf{E}\left\{\mathbf{r}\mathbf{d}^{\mathrm{H}}\right\}\} + \mathsf{E}\left\{\mathbf{d}\mathbf{d}^{\mathrm{H}}\right\}\right\}\\
=&\mathrm{tr}\left\{\mathbf{G}_m^{\mathrm{H}}(\mathbf{A}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I})\mathbf{G}_m - 2\Re\{\mathbf{G}_m^{\mathrm{H}}\mathbf{A}\} + \mathbf{I}\right\}.
\end{aligned}$$

Let $\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}}$ be the eigenvalue decomposition of $\mathbf{R}$. Then all matrices in the last equation has eigenvectors determined by $\mathbf{U}$, and we can therefore derive (14) as follows,

$$\begin{aligned}
&J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu},\alpha)\\
=\;&\mathrm{tr}\left\{\left(\mathbf{I} - \prod_{i=1}^{m}(\mathbf{I} - \mu_i(\mathbf{R}+\alpha\mathbf{I}))\right)(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\sigma^2\mathbf{I})\right.\\
&\times(\mathbf{R}+\alpha\mathbf{I})^{-1}\left(\mathbf{I} - \prod_{i=1}^{m}(\mathbf{I} - \mu_i(\mathbf{R}+\alpha\mathbf{I}))\right)^{\mathrm{H}}\\
&\left.-2\Re\left\{\left(\mathbf{I} - \prod_{i=1}^{m}(\mathbf{I} - \mu_i(\mathbf{R}+\alpha\mathbf{I}))\right)(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{R}\right\} + \mathbf{I}\right\}\\
=\;&\sum_{k=1}^{K}\frac{\lambda_k(\lambda_k+\sigma^2)}{(\lambda_k+\alpha)^2}\left|1 - \prod_{i=1}^{m}(1-\mu_i(\lambda_k+\alpha))\right|^2\\
&-2\Re\left\{\frac{\lambda_k}{(\lambda_k+\alpha)}\left(1 - \prod_{i=1}^{m}(1-\mu_i(\lambda_k+\alpha))\right)\right\} + 1\\
=\;&\sum_{k=1}^{K}\frac{\sigma^2}{\lambda_k+\sigma^2} + \sum_{k=1}^{K}\frac{\lambda_k(\lambda_k+\sigma^2)}{(\lambda_k+\alpha)^2}\\
&\cdot\left|\frac{\sigma^2-\alpha}{\lambda_k+\sigma^2} - \prod_{i=1}^{m}(1-\mu_i(\lambda_k+\alpha))\right|^2.
\end{aligned}$$

### B. Positive Semi-Definite $\mathbf{C}$

The matrix $\mathbf{C}$ is singular if and only if there exists an $\mathbf{x} \neq \mathbf{0}$ such that $\boldsymbol{\varphi}_k^{\top}\mathbf{x} = 0$ for $k = 1, 2, \cdots, K$, i.e.,

$$(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, ..., \boldsymbol{\varphi}_K)^{\top}\mathbf{x} = \mathbf{0}. \tag{26}$$

This constitutes $K$ equations in $m$ unknowns. If $m$ or more of the $K$ equations are linearly independent, then there is no non-zero solutions to (26). If we have less than $m$ linearly independent equations, there are multiple solutions. For $m = K$, (26) is a Vandermonde system [12] which is non-singular if and only if $\phi_i \neq \phi_j$ for all $1 \le i < j \le K$, or equivalently, $\mathbf{C}$ is non-singular if and only if $\mathbf{R}$ has $K$ distinct eigenvalues. Similarly, for $m \le K$, $\mathbf{C}$ is non-singular if and only if $\mathbf{R}$ has $m$ or more distinct eigenvalues.

### C. Real Optimal Weights for $\alpha = \sigma^2$

*Theorem 2:* The vector $\boldsymbol{\mu}$ that minimises

$$J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}) \triangleq J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}, \alpha = \sigma^2)$$

is real-valued.

*Proof:* Let $\boldsymbol{\mu} \in \mathbb{C}^m$ be the global minimum of $J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu})$ in $\boldsymbol{\mu}$. It is clear that

$$\begin{aligned}
J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}) &= \sum_{k=1}^{K}\frac{\lambda_k}{\phi_k}\prod_{i=1}^{m}|1-\mu_i\phi_k|^2\\
&= \sum_{k=1}^{K}\frac{\lambda_k}{\phi_k}\prod_{i=1}^{m}|1-\Re\{\mu_i\}\cdot\phi_k - j\Im\{\mu_i\}\cdot\phi_k|^2\\
&= \sum_{k=1}^{K}\frac{\lambda_k}{\phi_k}\prod_{i=1}^{m}\left[(1-\Re\{\mu_i\}\cdot\phi_k)^2 + (\Im\{\mu_i\}\cdot\phi_k)^2\right]\\
&\geq \sum_{k=1}^{K}\frac{\lambda_k}{\phi_k}\prod_{i=1}^{m}(1-\Re\{\mu_i\}\cdot\phi_k)^2\\
&= J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu}^{(r)}).
\end{aligned}\tag{27}$$

Therefore the excess MSE achieved by $\boldsymbol{\mu}$ is no less than that achieved by $\Re\{\boldsymbol{\mu}\}$. For $\boldsymbol{\mu}$ to be the global minimum, equality must hold in (27). It is obvious that $\Im\{\mu_i\} = 0$ for all $i$'s, i.e., $\boldsymbol{\mu} \in \mathbb{R}^m$. ■

If $\alpha \neq \sigma^2$, the above approach is not applicable and the optimal weights can be complex.

### D. Locating the Real Suboptimal Weights for $\alpha \neq \sigma^2$

*Theorem 3:* If the global minimum of $J_{\mathrm{ex}}^{(m)}(\boldsymbol{\mu},\alpha)$ in $\boldsymbol{\mu}$ is not real, its real-constraint minimum must be found in $\mathbb{B}_{\boldsymbol{\mu}} \triangleq \{\,\boldsymbol{\mu} \mid \exists\, 1 \le i < j \le m,\ \mu_i = \mu_j,\ \boldsymbol{\mu} \in \mathbb{R}^m\}$.

To prove this theorem, we need the following lemmas.

*Lemma 1:* Define the following set

$$\mathbb{D}_{\mathbf{x}} \triangleq \mathbf{T}(\mathbb{R}^m) = \{\,\mathbf{x} \mid \exists\,\boldsymbol{\mu} \in \mathbb{R}^m,\ \mathbf{x} = \mathbf{T}(\boldsymbol{\mu})\}\,.$$

If the global minimum of $J_{\mathrm{ex}}^{(m)}(\mathbf{x},\alpha)$ in $\mathbf{x}$ is not in $\mathbb{D}_{\mathbf{x}}$, its $\mathbb{D}_{\mathbf{x}}$-constraint minimum must be found in the boundary set of $\mathbb{D}_{\mathbf{x}}$.

*Proof:* As shown in Section V, $J_{\mathrm{ex}}^{(m)}(\mathbf{x},\alpha)$ has a unique minimum as $\hat{\mathbf{x}} = \mathbf{C}^{-1}\mathbf{p}$. If $\hat{\mathbf{x}} \notin \mathbb{D}_{\mathbf{x}}$, the constraint minimum in $\mathbb{D}_{\mathbf{x}}$ must be in the boundary set of $\mathbb{D}_{\mathbf{x}}$ since no local minimum exists within $\mathbb{D}_{\mathbf{x}}$. ■

*Lemma 2:* The boundary set of $\mathbb{D}_{\mathbf{x}}$ in $\mathbb{R}^m$, denoted by $\mathbb{B}_{\mathbf{x}}$, is determined by $\mathbb{B}_{\mathbf{x}} = \mathbf{T}(\mathbb{B}_{\boldsymbol{\mu}})$.

*Proof:* In the following we assume that all variables or series in $\mathbf{x}$ are real. Since $\mathbb{B}_{\mathbf{x}}$ is the boundary set of $\mathbb{D}_{\mathbf{x}}$,

$$\begin{gathered}
\forall\mathbf{x} \in \mathbb{B}_{\mathbf{x}},\ \forall\delta > 0,\ \exists\mathbf{x}^{(1)} \in \mathbb{D}_{\mathbf{x}},\ \mathbf{x}^{(2)} \notin \mathbb{D}_{\mathbf{x}},\\
\text{so that } \|\mathbf{x}^{(1)} - \mathbf{x}\| < \delta,\ \|\mathbf{x}^{(2)} - \mathbf{x}\| < \delta.
\end{gathered}\tag{28}$$

Therefore, for any $\mathbf{x} \in \mathbb{B}_{\mathbf{x}}$, two series exist,

$$\begin{gathered}
\exists\left\{\mathbf{x}_n^{(1)} \in \mathbb{D}_{\mathbf{x}}\right\} \text{ and } \left\{\mathbf{x}_n^{(2)} \notin \mathbb{D}_{\mathbf{x}}\right\},\\
\lim_{n\to\infty}\mathbf{x}_n^{(1)} = \lim_{n\to\infty}\mathbf{x}_n^{(2)} = \mathbf{x}.
\end{gathered}\tag{29}$$

By solving a polynomial with $\mathbf{x}$ as coefficients, we can find corresponding series in $\boldsymbol{\mu}$,

$$\begin{gathered}
\exists\left\{\boldsymbol{\mu}_n^{(1)} \in \mathbb{R}^m\right\} \text{ and } \left\{\boldsymbol{\mu}_n^{(2)} \notin \mathbb{R}^m\right\},\\
\lim_{n\to\infty}\mathbf{T}(\boldsymbol{\mu}_n^{(1)}) = \lim_{n\to\infty}\mathbf{T}(\boldsymbol{\mu}_n^{(2)}) = \mathbf{x}.
\end{gathered}\tag{30}$$

It is clear that $\left\{\boldsymbol{\mu}_n^{(1)}\right\}$ and $\left\{\boldsymbol{\mu}_n^{(2)}\right\}$ are bounded series, otherwise $\mathbf{T}(\boldsymbol{\mu}_n^{(1)})$ and $\mathbf{T}(\boldsymbol{\mu}_n^{(2)})$ will not converge. From theory of limitation, bounded series must have sub-series that converge. Therefore,

$$\begin{aligned} &\exists\left\{\boldsymbol{\mu}_{n_k}^{(1)} \in \mathbb{R}^m\right\},\ \lim_{k\to\infty} \boldsymbol{\mu}_{n_k}^{(1)} = \boldsymbol{\mu}^{(1)}; \\ &\exists\left\{\boldsymbol{\mu}_{n_k}^{(2)} \notin \mathbb{R}^m\right\},\ \lim_{k\to\infty} \boldsymbol{\mu}_{n_k}^{(2)} = \boldsymbol{\mu}^{(2)}. \end{aligned} \tag{31}$$

Since real vector series will converge only to a real vector, $\boldsymbol{\mu}^{(1)} \in \mathbb{R}^m$, which has corresponding $\mathbf{x} = \mathbf{T}(\boldsymbol{\mu}^{(1)}) \in \mathbb{D}_{\mathbf{x}}$. Since $\boldsymbol{\mu}^{(2)}$ is a solution to $\mathbf{T}(\boldsymbol{\mu}^{(2)}) = \mathbf{x}$, $\boldsymbol{\mu}^{(2)} \in \mathbb{R}^m$. $\mathbf{x} \in \mathbb{D}_{\mathbf{x}}$ implies that $\mathbb{B}_{\mathbf{x}} \subseteq \mathbb{D}_{\mathbf{x}}$. In the following we use the fact that $\boldsymbol{\mu}^{(2)} \in \mathbb{R}^m$ to prove that $\mathbb{B}_{\mathbf{x}} = \mathbf{T}(\mathbb{B}_{\boldsymbol{\mu}})$.

Rewriting $\boldsymbol{\mu}_{n_k}^{(2)}$ as $\boldsymbol{\nu}_k = \left(\nu_1^{(k)}, \nu_2^{(k)}, \cdots, \nu_m^{(k)}\right)^\top$ and $\boldsymbol{\mu}^{(2)}$ as $\boldsymbol{\nu}$, then

$$\forall\, k > 0,\ \boldsymbol{\nu}^{(k)} \notin \mathbb{R}^m,\ \boldsymbol{\nu} \in \mathbb{R}^m, \text{ and } \lim_{k\to\infty} \boldsymbol{\nu}^{(k)} = \boldsymbol{\nu}. \tag{32}$$

Since $\boldsymbol{\nu}^{(k)} \notin \mathbb{R}^m$ but $\mathbf{T}(\boldsymbol{\nu}^{(k)}) \in \mathbb{R}^m$, there must have at least one complex conjugate pair among $\nu_1^{(n)}, \nu_2^{(n)}, \cdots, \nu_m^{(n)}$, which are the $m$ roots of an $m^{\text{th}}$-order real coefficient polynomial. It is not difficult to deduce that there exist $i$ and $j$, $1 \le i < j \le m$, so that a sub-series of $\boldsymbol{\nu}^{(k)}$, denoted as $\left\{\boldsymbol{\xi}^{(l)} = \boldsymbol{\nu}^{(k_l)} \mid l = 1, 2, \cdots\right\}$, satisfies $\Re\{\xi_i^{(l)}\} = \Re\{\xi_j^{(l)}\}$ for all $l$. Therefore

$$\nu_i = \lim_{l\to\infty} \Re\{\xi_i^{(l)}\} = \lim_{l\to\infty} \Re\{\xi_j^{(l)}\} = \nu_j. \tag{33}$$

Equivalently, $\mu_i^{(2)} = \mu_j^{(2)}$, i.e., $\boldsymbol{\mu}^{(2)} \in \mathbb{B}_{\boldsymbol{\mu}}$. In conclusion, $\mathbb{B}_{\mathbf{x}} = \mathbf{T}(\mathbb{B}_{\boldsymbol{\mu}})$. ■

*Proof:* (Theorem 3) From Lemma 1, $\mathbb{D}_{\mathbf{x}}$ is the set of $\mathbf{x}$ that corresponds to real weights. If the global minimum of $J_{\text{ex}}^{(m)}(\boldsymbol{\mu}, \alpha)$ in $\boldsymbol{\mu}$, denoted by $\hat{\boldsymbol{\mu}}$, is not real, then the corresponding $\hat{\mathbf{x}} = \mathbf{T}(\hat{\boldsymbol{\mu}}) \notin \mathbb{D}_{\mathbf{x}}$. It follows that its real-constraint minimum, denoted as $\hat{\boldsymbol{\mu}}_r$, corresponds to the $\mathbb{D}_{\mathbf{x}}$-constraint minimum of $J_{\text{ex}}^{(m)}(\mathbf{x}, \alpha)$, which is found to be within $\mathbf{T}(\mathbb{B}_{\boldsymbol{\mu}})$ by Lemma 2. Consequently, $\hat{\boldsymbol{\mu}}_r \in \mathbb{B}_{\boldsymbol{\mu}}$. ■

## Acknowledgement

The authors would like to acknowledge the help and assistance provided by Mr. Ma Yao, Dr. Christopher Cheah, Dr. P. D. Alexander, Mr. H. Sugimoto, Dr. T. Oyama, and Mr. Y. Matsumoto.

## References

[1] A. J. Viterbi, *CDMA – Principles of Spread Spectrum Communication.* Addison–Wesley Publishing Company, 1995.

[2] F. Adachi, M. Sawahashi, and H. Suda, "Wideband DS-CDMA for next-generation mobile communications systems," *IEEE Communication Magazine*, vol. 36, pp. 56–69, Sept. 1998.

[3] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Communication Magazine*, vol. 34, pp. 124–136, 1996.

[4] S. Verdú, *Multiuser Detection.* Cambridge University Press, 1998.

[5] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 32, pp. 85–96, Jan. 1986.

[6] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.

[7] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detector for coherent multiuser communications," *IEEE J. Selected Areas Commun.*, vol. 8, pp. 683–690, May 1990.

[8] R. Lupas and S. Verdú, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, pp. 496–508, April 1990.

[9] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, April 1990.

[10] D. Divsalar, M. Simon, and D. Raphaeli, "Improved parallel interference cancellation for CDMA," *IEEE Trans. Commun.*, vol. 46, pp. 258–268, Feb. 1998.

[11] H. Elders-Boll, H. D. Schotten, and A. Busboom, "Efficient implementation of linear multiuser detectors for asynchronous CDMA systems by linear interference cancellation," *European Trans. on Telecommun.*, vol. 9, pp. 427–438, Sept./Oct. 1998.

[12] G. H. Golub and C. F. van Loan, *Matrix Computations.* The Johns Hopkins University Press, 2nd ed., 1989.

[13] L. K. Rasmussen, T. J. Lim, and A.-L. Johansson, "A matrix-algebraic approach to successive interference cancellation in CDMA," *IEEE Trans. Commun.*, vol. 48, pp. 145–151, Jan. 2000.

[14] S. Moshavi, E. G. Kanterakis, and D. L. Schilling, "Multistage linear receivers for DS-CDMA systems," *International Journal of Wireless Information Networks*, vol. 3, no. 1, pp. 1–17, 1996.

[15] T. Suzuki and Y. Takeuchi, "Real-time decorrelation scheme with multi-stage architecture for asynchronous DS/CDMA," *Technical Report of the IEICE*, vol. SST96-10, SAT96-24, RCS96-34, 1996. (in Japanese).

[16] D. Guo, L. K. Rasmussen, and T. J. Lim, "Linear parallel interference cancellation in long-code CDMA multiuser detection," *IEEE J. Selected Areas Commun.*, vol. 17, pp. 2074–2081, Dec. 1999.

[17] L. K. Rasmussen, D. Guo, T. J. Lim, and Y. Ma, "Aspects on linear parallel interference cancellation in CDMA," in *Proceedings 1998 IEEE International Symposium on Information Theory*, p. 37, MIT, Cambridge, MA USA, Aug. 1998.

[18] S. Haykin, *Adaptive Filter Theory.* Prentice Hall, 3rd ed., 1996.

[19] D. Guo, L. K. Rasmussen, S. Sun, T. J. Lim, and C. Cheah, "MMSE–based linear parallel interference cancellation in CDMA," in *Proceedings IEEE Fifth International Symposium on Spread Spectrum Techniques and Applications*, vol. 3, pp. 917–921, Sun City, South Africa, Sept. 1998.

[20] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inform. Theory*, vol. 43, pp. 858–871, May 1997.

[21] D. Guo, "Linear parallel interference cancellation in CDMA," M.Eng. thesis, National University of Singapore, 1998.